#6 RS

10-25-01

Attorney Docket: 3036/49962
PATENT

09/864880

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: ANDREW REEVE ET AL.

Serial No.: NOT YET ASSIGNED

Filed: May 25, 2001

Title: IMPROVEMENTS IN OR RELATING TO PACKET SWITCHES

CLAIM FOR PRIORITY UNDER 35 U.S.C. §119

Box PATENT APPLICATION
Commissioner for Patents
Washington, D.C. 20231

May 25, 2001

Sir:

The benefit of the filing date of prior foreign application No. 0012600.3, filed in Great Britain on 25 May 2000 and No. 0024455.8, filed in Great Britain on 6 October 2000 is hereby requested and the right of priority under 35 U.S.C. §119 is hereby claimed.

In support of this claim, filed herewith is a certified copy of the original foreign application.

Respectfully submitted,

/39085

Gary R. Edwards
Registration No. 31,824

for

CROWELL & MORING, LLP
P.O. Box 14300
Washington, DC 20044-4300
Telephone No.: (202) 628-8800
Facsimile No.: (202) 628-8844
GRE:kms

THIS PAGE BLANK (USPTO)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

INVESTOR IN PEOPLE

CERTIFIED COPY OF PRIORITY DOCUMENT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed M Cooki

Dated 15 May 2001

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

25MAY00 E539605-1 D00393
P01/7700 0.00-0012600.3

THE PATENT OFFICE
L
25 MAY 2000
NEWPORT

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

2 5 MAY 2000

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1.	Your reference	2000P04862/GB/R76/DA/cs
2.	Patent application number *(The Patent Office will fill in this part)*	**0012600.3**
3.	Full name, address and postcode of the or of each applicant *(underline all surnames)*	Roke Manor Research Limited Roke Manor Old Salisbury Lane Romsey Hampshire. SO51 0ZN
	Patents ADP number *(if you know it)*	5615455005
	If the applicant is a corporate body, give the country/state of its incorporation	United Kingdom
4.	Title of the invention	IMPROVEMENTS IN OR RELATING TO CROSS-BAR SWITCHES
5.	Name of your agent *(if you have one)*	MARGARET MACKETT
	"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	Siemens Shared Services Limited Intellectual Property Department Siemens House, Oldbury Bracknell, Berkshire RG12 8FZ United Kingdom
	Patents ADP number *(if you know it)*	7761000002.

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:* — YES

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	7
Claim *(s)*	0
Abstract	0
Drawing *(s)*	5 x 5 M

10. If you are also filing any of the following, state how many against each item.

Priority documents	0
Translations of priority documents	0
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	0
Request for preliminary examination and search *(Patents Form 9/77)*	0
Request for substantive examination *(Patents Form 10/77)*	0
Any other documents *(please specify)*	0

11. I/We request the grant of a patent on the basis of this applicatio

Signature M D Mackett Date 24/05/2000

Margaret Mackett

12. Name and daytime telephone number of person to contact in the United Kingdom

Margaret Mackett - 01344 396808

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publica or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. Yo will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in th United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first get written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 50050*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separat sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

IMPROVEMENTS IN OR RELATING TO CROSS-BAR SWITCHES

The present invention relates to improvements in or relating to cross-bar switches.

Traffic volume in the Internet is growing exponentially, almost doubling every 3 months. The current capacity of IP Routers is insufficient to meet this demand and hence there is a market opportunity for products that can route IP traffic at extremely large aggregate bandwidths in the order of several Terabit/s. Such routers are termed "Terabit Routers".

Two important trends are also evident. First operators are consolidating all traffic onto a single IP back-bone. Second IP is increasingly required to support real-time and multimedia traffic. This means that the next generation of routers must also support Quality of Service (QoS). In particular they must support low bounded delay for real-time traffic.

Terabit Routers will require a scalable high capacity communications path between its line functions. One technique that may be employed is a cell based cross-bar, described below. However, the effectiveness of this technique with respect to the router's efficiency and ability to support the quality of service of the communications it supports depends on the cross-bar configuration algorithm employed.

This invention described here is a cross-bar configuration algorithm which provides very high levels of efficiency and support for quality of service. It has application in any switch or router such as IP Routers, ATM switches or MPLS Label Switch Routers.

For a better understanding of the present invention, reference will now be made, by way of example only, to the accompanying drawings in which:-

Figure 1 illustrates a Terabit router architecture;

Figure 2 illustrates a cross-bar controller;

Figure 3 illustrates a rate matrix (8x8);

Figure 4 illustrates an example cross-bar configuration matrix (8x8); and

Figure 5 illustrates a 'find_config' procedure.

Cross-bar architectures are already in use in routers and switches, but they typically suffer from

1. the inability to achieve 100% utilisation of the cross-bar without incurring very long delays
2. the inability to guarantee upper bounds for delay as required to support real-time traffic.

This section describes a terabit router architecture and explains what is meant by a cell based cross-bar.

Figure 1 illustrates a Terabit router architecture 100 in which packets arrive at ingress forwarders 102, 104, 106 via their input port(s) and are routed across a cross-bar 110 to a correct egress forwarder 120 which transmits them across its output port(s). Each ingress forwarder maintains a separate packet queue for each egress forwarder.

For the purpose of this discussion, a cell based cross-bar is characterised as follows:

1. Each ingress line function may be connected to any egress line functions.

2. Each ingress line function may only be connected to one egress line function at a time.

3. Each egress line function may only be connected to one ingress line function at a time.

4. All ingresses transmit in parallel across the cross-bar.

5. Data is transmitted across the cross-bar in small fixed sized cells. An example cell size is 64 octets.

6. Cell transmission is synchronised across all the ingress line functions. This means that for each cell cycle, each ingress line function starts transmitting the next cell at the same time.

7. The cross-bar is reconfigured at the end of every cell cycle.

The invention is a cross-bar unicast cell scheduling algorithm for which the following requirements must be satisfied:-

1. **starvation free.** No packet queue with non-zero occupancy must ever be starved of bandwidth across the cross-bar.

2. **fair**. The algorithm must be fair in its allocation of bandwidth and priority to packet queue.

3. **provide configured rates.** The algorithm must implement rates configured for each ingress packet queue q_{jk}. The rates are calculated to ensure the QoS of traffic streams.

4. **bounded low delay.** The algorithm must provide bounded low delay across the cross-bar. Note that the overall delay across the whole router will also be determined by the traffic management functions implemented in the LICs.

5. **efficiency.** The algorithm must support 100% utilisation of the cross-bar without loss of performance.

An embodiment of a cross-bar controller arrangement 200 in accordance with the present invention is depicted in Figure 2.

The cross-bar controller arrangement 200 comprises a cross-bar 202 which is controlled by a cross-bar controller 204 which in turn is controlled by a bandwidth controller 206. The bandwidth controller 206 is responsible

for efficient allocation of the bandwidth across the cross-bar 202, and calculates the rates that each ingress forwarder 210, 212, 214, 216 must transmit to each egress forwarder 220, 222, 224, 226. This is the same as the rate at which data must be transmitted from each packet queue. The means by which these rates are calculated is beyond the scope of this description.

The bandwidth controller 206 transmits the rates to the cross-bar controller 204 which is responsible for efficient scheduling of data across the cross-bar 202 whilst maintaining the rates calculated by the bandwidth controller 206. It is responsible for calculating the following information at the end of each cell cycle.

1. To each ingress forwarder it transmits the identity of the next packet queue from which to transmit. Recall that each ingress forwarder maintains a separate queue of packets for each egress forwarder.

2. To the cross-bar it transmits the next cross-bar configuration.

A description of how the cross-bar controller 204 determines this information follows:-

The rates can be represented using a matrix as depicted in Figure 3, which provides the example of an 8×8 router. Let us call this matrix R, with elements r_{jk}, such that r_{jk} is the rate from ingress forwarder j to egress forwarder k. The rate unit employed is cells per unit time.

Let F be the rate at which an ingress forwarder 210, 212, 214, 216 receives data from its input port and transmits across the cross-bar 202. F is also the rate at which an egress forwarder 220, 222, 224, 226 receives data from the cross-bar port and transmits across its output port.

Let N be the number of ingress forwarders 210, 212, 214, 216. N is also the number of egress forwarders 220, 222, 224, 226.

Then:-

$$\sum_{k=1}^{N} r_{jk} \leq F \tag{1}$$

$$\sum_{k=1}^{N}\sum_{j=1}^{N} r_{jk} = N \tag{2}$$

$$\sum_{k=1}^{N}\sum_{j=1}^{N} r_{jk} = N \tag{3}$$

Herein, a matrix with the property that the column and row sums are all equal shall be called a sum-perfect square. This is loose use of the term since a mathematical sum-perfect square has stronger properties, including that no entry has the same value, which is not helpful here.

The configuration of the cross-bar 202 can be depicted by a matrix such as that shown in Figure 4. Each entry in the matrix may take the value 0 or 1. An entry with the value 1 indicates that the cross-bar 202 is configured for transmission of cell from the corresponding ingress (row index) to the corresponding egress (column index). A value of 1 indicates across-connect between the ingress and egress. A value of 0 indicates absence of a cross-connect.

For unicast operation the cross-bar configuration matrix has the following properties:-

1. Each row contains exactly one non-zero entry. (Each ingress transmits a cell to exactly one egress.)

2. Each column contains exactly one non-zero entry. (Each egress can receive a cell from one ingress only.)

Note that these properties mean that the matrix forms a sum-perfect square where the row/column sum is 1. Let us call this matrix X.

The algorithm operates as follows.

1. The rate matrix R is re-calculated periodically.
2. The rate is maintained over the length of each period.

3. At the beginning of each period, calculate a matrix containing the number of cells that must be transmitted from each of the packet queues q_{jk} during the period to achieve its rate.

4. For each cell slot in the period, find a configuration which matches the cell count matrix by only servicing queues with non-zero cell counts.

5. Decrement by 1 the cell counts of each queue serviced.

6. Repeat from 3. until the end of the period at which all cell counts will be zero.

This is expanded below:-

Let P be the length of the period. This is chosen to be a whole number of cell cycles and to be of magnitude no greater than the maximum tolerated delay for real-time traffic.

Let T be the cell slot number in the period P, such $T = 0$ is the first slot and $T = P - 1$ is the last.

Let C^T be the cell count matrix at the beginning of cell slot T. Then C^0 is calculated as follows:-

$$C^0 = PR \tag{4}$$

The rate matrix R is calculated such that C^0 contains integral values. How this is done is beyond the scope of this description.

Note that since R is semi-magic, then so is C^0.

For each T, calculate a cross-bar configuration matrix X^T such that

$$x^T_{jk} = 1 \Rightarrow c^T_{jk} > 0 \tag{5}$$

Since X^T provides a cell send opportunity for each q_{jk} such that $x^T_{jk} = 1$, C^T is calculated as follows:-

$$C^{T+1} = C^T - X^T \tag{6}$$

Since C^0 is sum-perfect and X^T is sum-perfect for all T, then it is easy to prove that C^T is sum-perfect for all T. In particular C^P is sum-perfect with sum 0 so that all elements are zero and all queues have been given their full complement of cell send opportunities.

An example algorithm for finding a configuration matrix X^T matching cell count matrix C^T is presented below. The main routine calls the 'find_config' routine for the number of time slots in period P. 'find_config' is a recursive heuristic algorithm.

The main routine is:

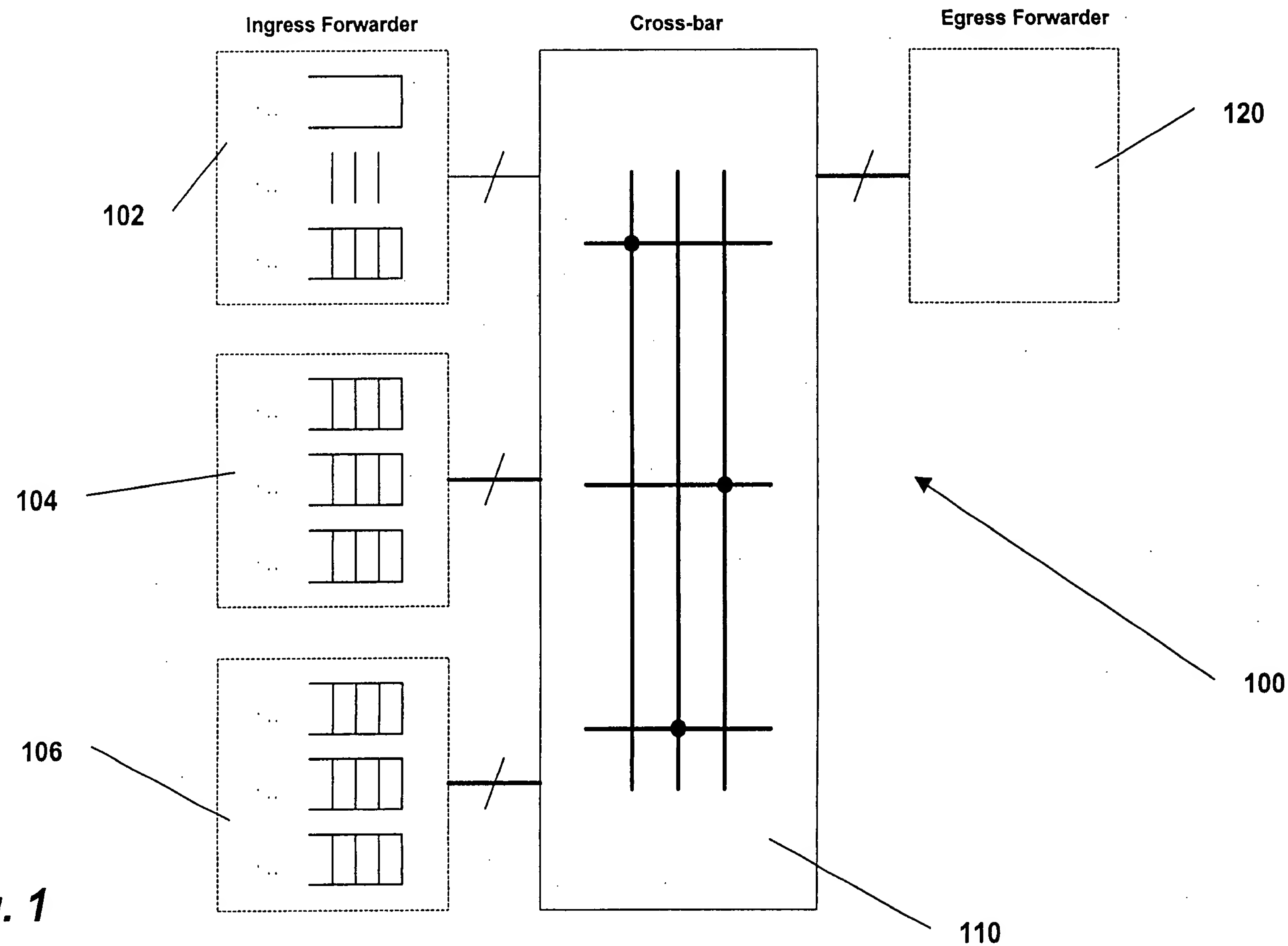
for slot = 0 **to** period_P **do**

begin

  if not find_config(ingress = 0) **then abort**;

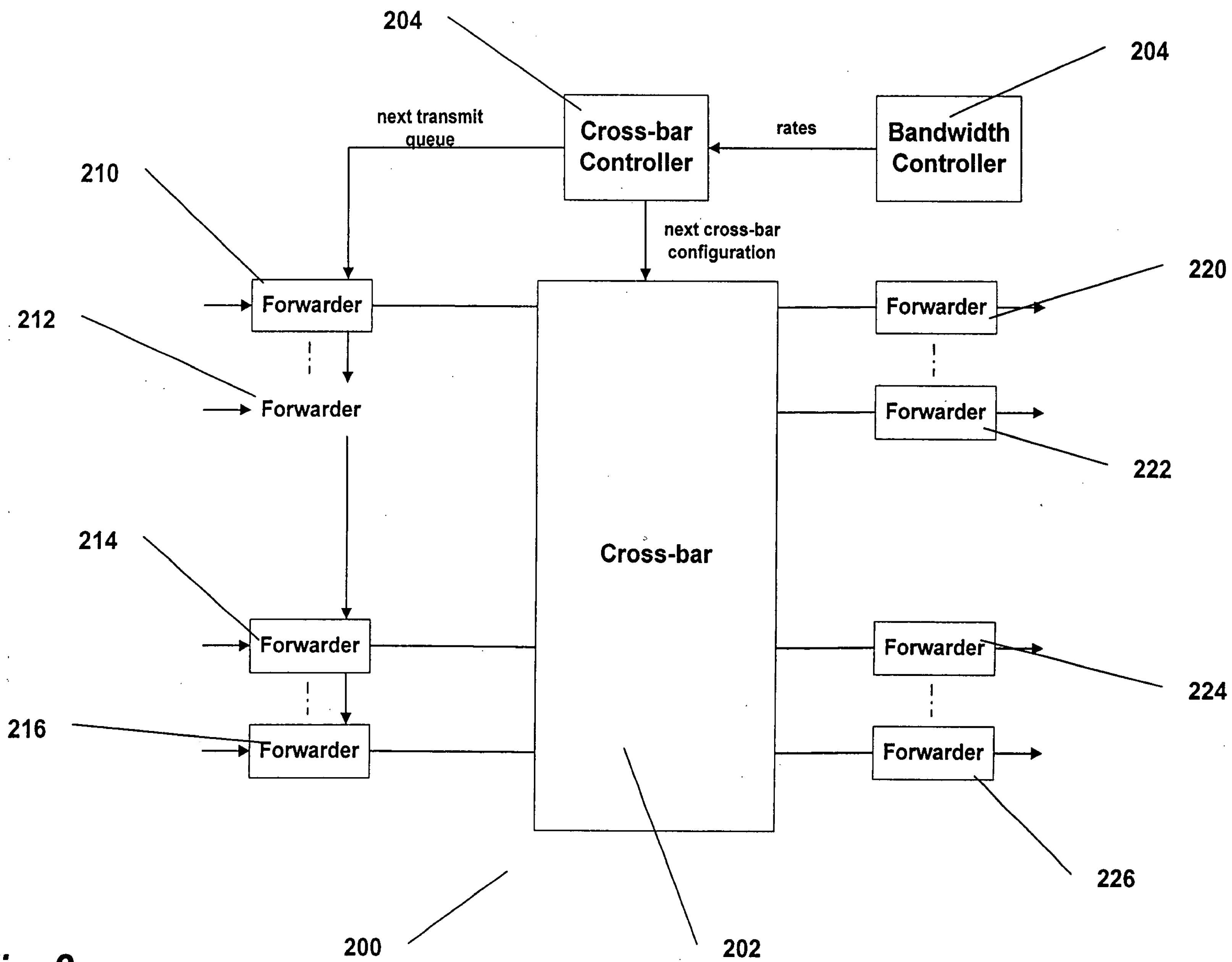
end

The 'find_config' routine is shown in Figure 5.



Ingress Forwarder
Cross-bar
Egress Forwarder
102
104
106
110
120
100

Fig. 1



204
Cross-bar Controller
next transmit queue
rates
204
Bandwidth Controller
next cross-bar configuration
210
Forwarder
212
Forwarder
214
Forwarder
216
Forwarder
Cross-bar
220
Forwarder
Forwarder
222
Forwarder
224
Forwarder
226
200
202

Fig. 2

THIS PAGE BLANK (USPTO)

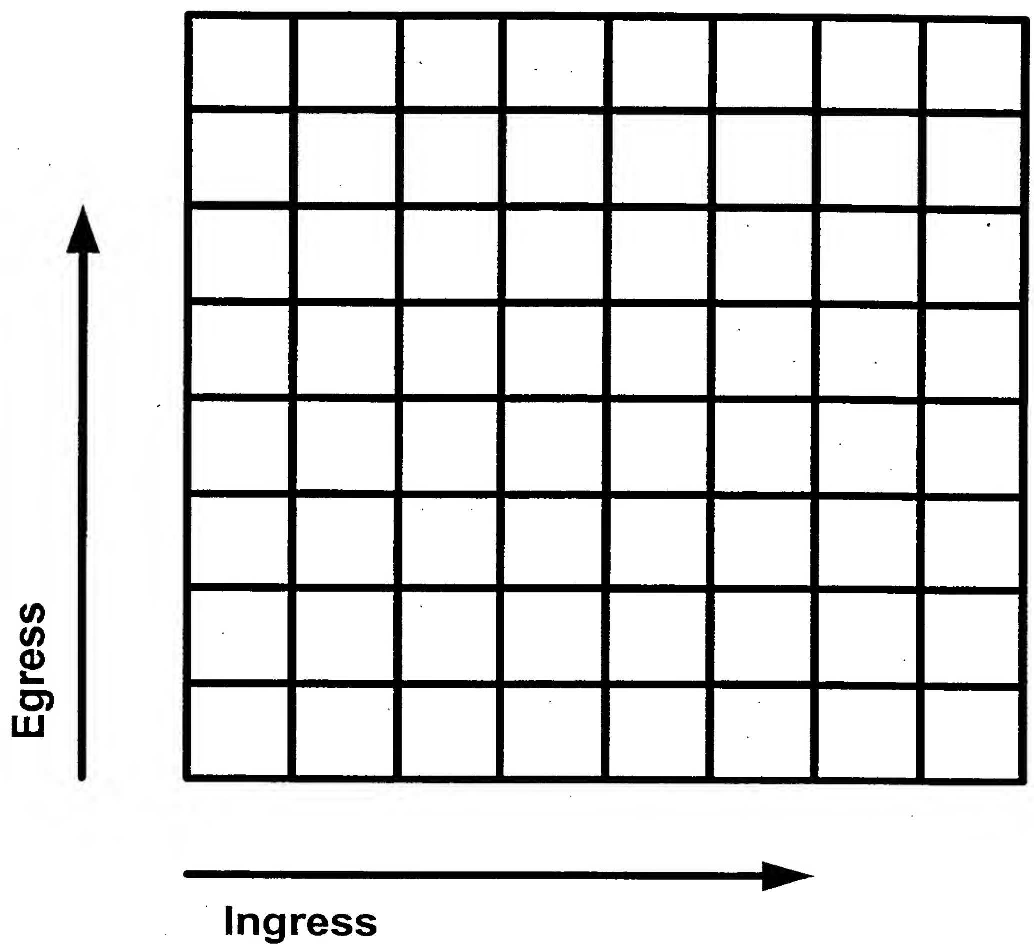


Fig. 3

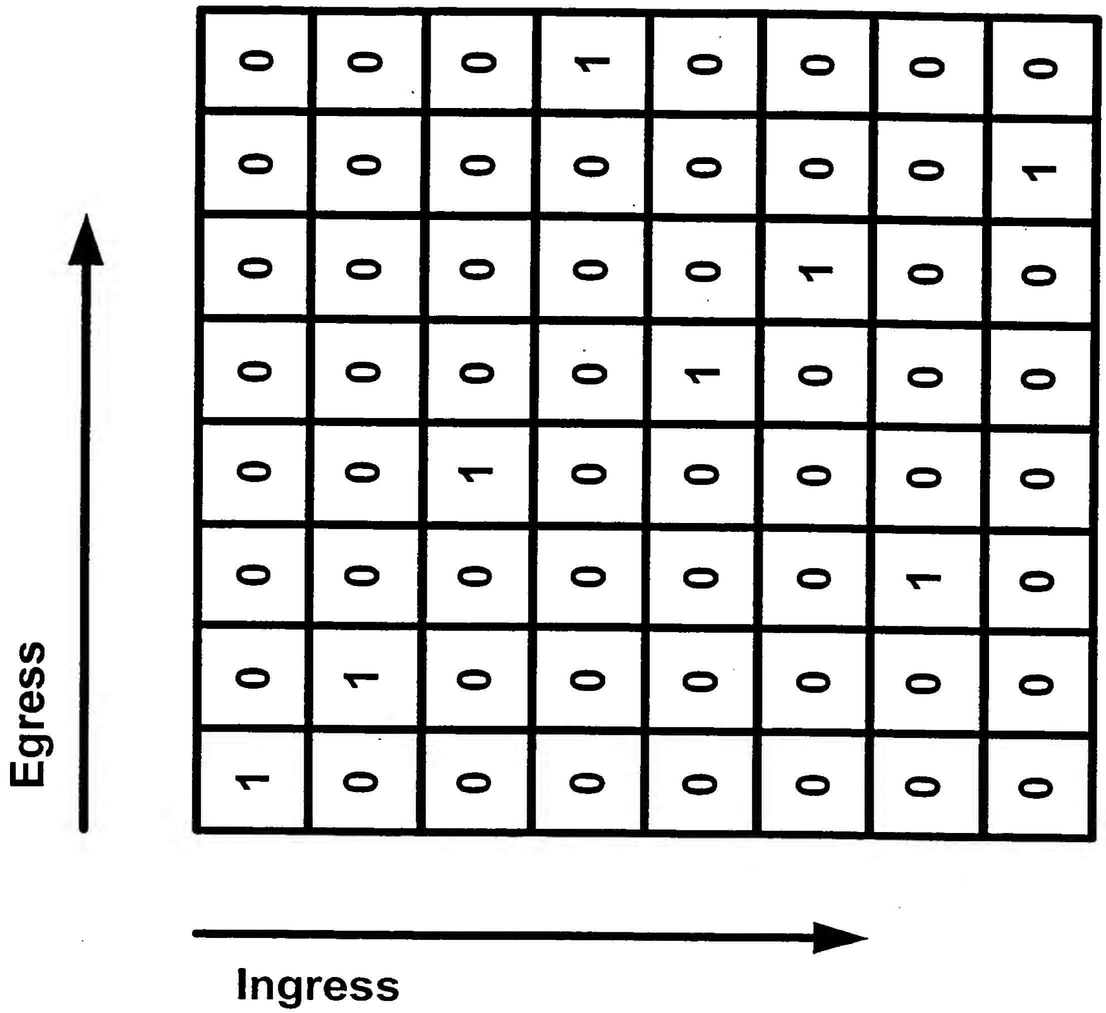


Fig. 4

```
find_config(ingress) returns boolean
begin

    for all egress ports do
    begin

        if ( (cell_count[ingress, egress] > 0) and picked[egress] == false) then
        begin
            picked[egress] = true;
            if ( (ingress == last ingress) or find_config(ingress+1)) then
                return true;
            else
                picked[egress] = false;
        end

    end

    return false;          // no match configuration found
end
```

Fig. 5